Probabilistic Forecasts of Wind Speed using the Bootstrapped Markov Regime Switching Model

William Yoo Weimin and Yulia R. Gel

Technical Report

Department of Statistics and Actuarial Science, University of Waterloo

March, 2010

Abstract

With a recent surge of different alternative fuels, wind power has proven to be one of the most attractive option. In order to harvest its power, accurate and timely short-term forecasts of wind speeds are needed. We introduce the Bootstrapped Markov Regime Switching (BMRS) model to produce fully probabilistic 2-hour ahead forecasts of wind speed. Our model building strategy consists of 2 elements: The first is to explore the possibility of producing competitive Bayesian probabilistic forecasts using only on-site wind information as opposed to spatial-temporal approaches. Secondly, we hope to eliminate the need to specify beforehand a parametric predictive distribution and adopt bootstrap methodologies to produce probabilistic forecasts. By coding 90° to 270° clockwise as West and 90° to 270° anticlockwise as East, we assume wind speed is driven by two different autoregressive processes, each corresponding to a different regime of wind direction (East or West), and the transition from one regime to another is governed by a two state Markov chain. In this paper, we introduce the circular Markovian local bootstrap technique to model future state evolution. The parameters are estimated using a modified Gibbs sampler, while the final distribution-free predictive intervals for future wind speeds are developed using sieve bootstrap. The proposed method is applied to real life wind speed data in a case study.

KEY WORDS: Markov regime switching model, Markov chain Monte Carlo, Gibbs sampler, Probabilistic forecast, Markovian local bootstrap, Sieve bootstrap

1 Introduction

In the midst of a recent proliferation of different alternative fuels, wind power has proven to be an attractive option. In the past decade or so, wind power has been the fastest growing renewable power source around the globe percentage-wise, with an annual average growth rate exceeding 30%. Today, wind energy continues to attract the interest of the public and the government because it is a highly reliable and environmental friendly energy source. Interested readers can refer to the American Wind Energy Association website for the most recent developments in wind energy. However, worldwide utilization of wind power is still heavily concentrated on the European continent and the United States. For example, wind power has successfully supplied 20% of Denmark's energy needs and is projected to meet around 6% of U.S's energy needs by 2020. For the rest of the world, the notion of wind energy is still relatively new. The main hurdle that prevents the proliferation of wind energy is the perceived difficulties involved in producing accurate and timely short term forecast of wind speed. In many cases, spatial-temporal models are used where several off-site wind observations are incorporated into the forecasting process in order to improve the quality of the forecasts. In many areas around the globe, especially in the developing countries, existing wind speed records are often sparse and incomplete due to the inhibiting cost of wind tower construction and its subsequent maintenance. Lacking credible off-site observations, these models will be ineffective when applied to these situations. This paper proposes using the Bootstrapped Markov Regime Switching model (BMRS) to produce probabilistic forecasts of wind speed utilizing only on-site wind speed and wind direction data. Here, wind direction is used as an auxiliary data to improve forecasting performance. Since wind speed and direction data are almost always collected together, it is hoped that this proposed model will find wide application in the wind energy industry.

To use this energy resource, we need to find a way to identify potential geographical areas that experience consistently high wind speed. As wind power is a non-linear function of wind speed, areas with high wind speed will prove to be the best site for wind turbine or wind farm construction. Thus, we need to produce short-term forecasts of wind speed at a chosen site to evaluate its potential to produce wind energy. This paper attempts to introduce the usage of Markov regime switching model to produce fully probabilistic forecasts of wind speed data.

There exists various methodologies and techniques for short term wind speed prediction, ranging from persistence-based forecasts to state-of-the-art numerical weather prediction models. Being one of the earliest methods introduced, persistence forecast models use the current observation as point forecast for all future periods. These models are improved upon by using univariate autoregressive processes to model wind speed observations (Brown et al., 1984). A recent breakthrough in this literature is the introduction of the Regime-Switching Space-Time Diurnal model (RSTD) by Gneiting et al. (2006). This model first identifies atmospheric regimes, in this case, wind directions at any potential wind energy sites and then fits predictive truncated normal distribution to each regime to produce probabilistic forecasts. The separation of regimes are in turn determined by up-wind direction records. Here, the mean and variance parameters of the forecast distribution are linear functions of on-site and off-site present and past observations. Hering and Genton (2010) then extended this idea by incorporating wind direction as additional regressors and further introduced the notion of converting wind data into two-dimensional Cartesian wind vector to model wind speed dynamics.

In this paper, we attempt to model wind speed data using only on-site information. Also, we will use the corresponding wind direction as an auxiliary variable to improve forecast performance. To achieve these objectives, we consider using a Markov regime switching model to take into account the non-linear dynamics of wind speed data. The Markov regime switching model was initially introduced by Hamilton (1988, 1989) to model a transition between different states in the economy. Subsequently, this model has been widely analyzed and applied to various time series data. For example, McCulloch and Tsay (1994) used this model to analyze the U.S quarterly real GNP growth rate and introduced the idea of using Markov chain Monte Carlo method for parameter estimation. In addition, we explore the possibility of producing probabilistic forecasts of wind speed without making any distribution assumptions by sieve bootstrapping historical residuals. There are numerous papers on the application of sieve bootstrap to time series and forecasting and we will use the version introduced by Alonso et al. (2002) for the purpose of our paper. Furthermore, the idea of applying bootstrap methodologies to the context of a Markov regime switching model was discussed by Psaradakis (1998), where he introduced the idea of using bootstrap resampling techniques to evaluate the adequacy of a Markov switching autoregressive model. Paparoditis and Politis (2001, 2002) then introduced the Markovian local bootstrap, which is an algorithm to generate bootstrap samples from a Markov process. A modified version of this technique will be used in this paper to simulate possible sample paths of the underlying Markov process.

The rest of the paper is organized as follows. In Section 3, we describe the data used, other main wind predictive models, and introduce the Bootstrapped Markov regime switching model. Here, we assume that wind speed is driven by two different autoregressive processes with each corresponding to a different regime of wind direction, culminating in a detailed explanation of the parameter estimation procedures. In this model setting, to predict future wind speed, we need to first predict future wind direction. To achieve this, we will use a modified version of the Markovian local bootstrap technique to model future state evolution. We then describe the steps taken to produce probabilistic forecasts of wind speed using sieve bootstrap methodologies by Alonso et al. (2002). In Section 4, we illustrate this proposed model using actual wind speed and direction data. Here, we also describe investigations conducted in order to study the structure of the data and to determine the order of the autoregressive processes used in the proposed model. In Section 5, we assess the predictive performance of the proposed model using various performance measures that are available in the literature. The paper then ends with Section 6, where we will discuss possible improvements and shortcomings of the proposed model.

2 Wind speed prediction models

2.1 Data description

The wind speed and wind direction data used for this study are obtained from Oregon State University's Energy Resources Research Laboratory. These data were collected by meteorological towers around the vicinity at the U.S. Pacific Northwest. In this paper, we focus our attention on wind speed and wind direction observations collected from the wind towers at Vansycle, Kennewick and Goodnoe Hills near northeastern

Oregon. These data are available in 2 forms: the first or the original version consists of data collected every 10 minute, while the second version is obtained by aggregating and averaging the original 10-minute data to get hourly data series. Since our purpose is to produce 2-hour ahead forecasts, we will use the latter version. The data from all 3 towers are simultaneously available starting from August 2002. Wind speed is measured in miles per hour (mph) and will be converted to meters per second (ms^{-1}) for forecasting and comparison purposes, while wind directions are recorded in degrees. Readers are encouraged to refer to the original paper by Gneiting et al. (2006) to get a full description of the data and a more detailed profile of the wind towers mentioned. Figure 1 below shows the locations of Vansycle, Kennewick and Goodnoe Hills near the Washington-Oregon border. In the map, Kennewick lies 39 km northwest of Vansycle and Goodnoe Hills lies 146 km west of Vansycle.

Figure 1: Locations of Vansycle, Kennewick and Goodnoe Hills in the U.S Pacific Northwest.

2.2 Regime-Switching Space-Time Diurnal Model

To begin, we will give a brief introduction to the Regime-Switching Space-Time Diurnal (RSTD) model introduced by Gneiting et al. (2006). This is then followed by a very brief review of the Trigonometric Direction Diurnal (TDD) and the Bivariate

Skew-T (BST) models, both introduced by Hering and Genton (2010). These models were first mentioned in the introduction but will be described in more detail for clarity and also to serve as a motivation to our proposed model. It should be noted that readers are encouraged to refer to the original papers for a more complete model description.

Denote V_t, K_t and G_t to be the wind speed data at time t for Vansycle, Kennewick and Goodnoe Hills respectively. In the RSTD model, the 2 different regimes are defined by the corresponding wind direction at Goodnoe Hills, e.g. if Goodnoe Hills experience westerly wind 2 hours beforehand, then the current regime at Vansycle is westerly and vice versa. At the westerly regime, we fit 2 pairs of sine and cosine functions to remove the diurnal pattern to get residual series V_t^r, G_t^r, and K_t^r. The 2-hour ahead probabilistic forecast of hourly Vansycle wind speed, V_{t+2} is assumed to have a truncated normal distribution with mean μ_{t+2} and standard deviation σ_{t+2}. In other words, $V_{t+2} \sim N^+(\mu_{t+2}, \sigma_{t+2}^2)$, where $N^+(\cdot,\cdot)$ represents the truncated normal distribution. Here, the mean and standard deviation parameters are linear combinations of present and past values at all 3 towers. In the westerly regime, the predictive mean equation is given by

$$\mu_{t+2} = D_{t+2} + \mu_{t+2}^r, \tag{2.1}$$

where D_{t+2} is the fitted diurnal pattern at Vansycle and is given by

$$d_0+d_1\sin\left(\frac{2\pi(t+2)}{24}\right)+d_2\cos\left(\frac{2\pi(t+2)}{24}\right)+d_3\sin\left(\frac{4\pi(t+2)}{24}\right)+d_4\cos\left(\frac{4\pi(t+2)}{24}\right),$$

while μ_{t+2}^r is given by

$$\mu_{t+2}^r = a_0 + a_1V_t^r + a_2V_{t-1}^r + a_3K_t^r + a_4K_{t-1}^r + a5G_t^r. \tag{2.2}$$

On the other hand, in the easterly regime, Gneiting et al. (2006) do not fit trigonometric functions to remove the diurnal component and the predictive mean equation is

$$\mu_{t+2} = a_0 + a_1V_t + a_2K_t \tag{2.3}$$

In addition, the RSTD attempts to take conditional heteroscedasticity into account by modeling σ_{t+2} as a linear function of the volatility value v_t, given by

$$\sigma_{t+2} = b_0 + b_1v_t. \tag{2.4}$$

In the westerly regime, the volatility value is given by

$$v_t = \left(\frac{1}{6} \sum_{i=0}^{1} ((V^r_{t-i} - V^r_{t-i-1})^2 + (K^r_{t-i} - K^r_{t-i-1})^2 + (G^r_{t-i} - G^r_{t-i-1})^2) \right)^{1/2}. \quad (2.5)$$

For the volatility value in the easterly regime, the residual series in equation (2.5) are replaced by the original series. Then, the parameters from equations (2.1) to (2.5) are estimated numerically by minimizing the Continuous Ranked Probability Score (CRPS) for a truncated normal distribution. Full details of the estimation step and the constraints used for minimization can be found in the original paper by Gneiting et al. (2006). For this model, the point forecast for V_{t+2} is the mean μ^+_{t+2} and can be shown to take the form of

$$\mu^+_{t+2} = \mu_{t+2} + \sigma_{t+2}\phi\left(\frac{\mu_{t+2}}{\sigma_{t+2}}\right)/\Phi\left(\frac{\mu_{t+2}}{\sigma_{t+2}}\right), \quad (2.6)$$

and the α-quantile is given by

$$Q^+_{\alpha,t+2} = \mu_{t+2} + \sigma_{t+2}\Phi^{-1}[\alpha + (1-\alpha)\Phi(-\mu_{t+2}/\sigma_{t+2})]. \quad (2.7)$$

2.3 Trigonometric Direction Diurnal and Bivariate Skew-T Models

The TDD model by Hering and Genton (2010) serves as an extension or generalization to the RSTD model. This approach eliminates the need to define different regimes by incorporating wind direction directly into the predictive mean equations. First, in addition to removing diurnal trend for wind speed as mentioned above, they did the same for wind direction. Then they include the sine and cosine of these de-trended wind directions into equation (2.2) using Bayesian Information Criterion (BIC) to produce a single predictive mean equation without any separation of regimes The BST method approaches this problem from a different perspective. In this model, they first convert wind speed and direction into Cartesian components with $x = r\cos(\theta)$ and $y = r\sin(\theta)$ where r is wind speed and θ is wind direction. Then, denote $\boldsymbol{V}_t = (V_{t,x}, V_{t,y})'$ to be the wind vector at time t, where $V_{t,x}$ is the east-west component and $V_{t,y}$ is the north-south component. Also, let $\boldsymbol{K}_t$ and $\boldsymbol{G}_t$ be the similar wind vectors at Kennewick and Goodnoe Hills respectively. Then, we remove the diurnal components and standardize these wind vectors. The resulting standardized residual wind vector series are then used to model 2-step ahead Vansycle wind speed by replacing the scalar wind speed series V_t, K_t and G_t in (2.2) with wind vectors, $\boldsymbol{V}_t$, $\boldsymbol{K}_t$ and $\boldsymbol{G}_t$. Also, the scalar coefficients are replaced by the corresponding vectors or matrices and a bivariate skew-t distributed error random variable is added as a last term in (2.2). Besides that, equation (2.3) is modified slightly to accommodate these changes. Here, the main difference is that the predictive equation no longer takes the form of a truncated normal but a bivariate skew-t distribution.

The parametric specification of a predictive distribution may be a convenient and sometimes efficient way to produce probabilistic forecasts. However, the accuracy of these approach depend heavily on the validity of the parametric models used to model the error structure. If the assumed predictive distribution does not closely match the underlying true distribution, then the probabilistic forecasts produced will be biased and inaccurate. Thus, a lot of preliminary investigations have to be done in order to come up with a reasonable predictive distribution. Another approach to this problem is to use non-parametric approaches to produce these forecasts. One popular method would be bootstrap and we will use this concept repeatedly in our model building process. Also, as mentioned in the introduction on the need to rely only on on-site wind information, we restrict ourself to use only on-site wind data. Thus, our model building strategy consists of 2 elements: The first is to explore the possibility of producing competitive probabilistic forecasts using only on-site wind information as opposed to the spatial-temporal approaches adopted by both models described above. Secondly, we hope to eliminate the need to specify beforehand a parametric predictive distribution and adopt bootstrap methodologies to produce probabilistic forecasts. To achieve this, we use ideas from the RSTD and TDD models and also experimented with various possible combination of models that are currently available in the literature. We find out that the novel combination of Markov regime switching model, sieve bootstrap and Markovian local bootstrap techniques provide us a way to achieve our stated model aims and strategies.

2.4 Bootstrapped Markov Regime Switching Model

2.4.1 Model description

Let x_t denote wind speed observations measured at time t. In the context of this paper, x_t can be any of wind speed series V_t, K_t or G_t as defined in the previous section. As will be shown in the case study, we will concentrate on modeling wind speed at Vansycle, V_t, but for generality, we will use the notation x_t to represent any wind speed time series. Therefore, we can model the dynamics of x_t using a general Markov regime switching model as follows:

$$x_t = \begin{cases} c_1 + \sum_{i=1}^{p} \phi_{1,i} x_{t-i} + \epsilon_{1t} & \text{if } S_t = 1, \\ c_2 + \sum_{i=1}^{p} \phi_{2,1} x_{t-i} + \epsilon_{2t} & \text{if } S_t = 2, \end{cases} \tag{2.8}$$

where c_1 and c_2 are constants that represent the mean level for each regime. The variable p is the order of the autoregressive process and is assumed to be equal for both regimes. The value of p can be estimated by observing the autocorrelation function (ACF) plot of the wind speed data. An alternative approach would be to select the value of p that minimizes a certain information criterion such as Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC). Here, the parameters $\phi_{1,i}$ and $\phi_{2,i}$ are the autoregressive coefficients that satisfy the necessary condition for stationarity, i.e. the solutions of characteristic equations for $\phi_{1,i}$ and $\phi_{2,i}$ lie outside of the unit circle. The innovation series $\{\epsilon_{1t}\}$ and $\{\epsilon_{2t}\}$ are assumed to be i.i.d. random variables with means 0 and finite variances σ_1^2, σ_2^2 respectively and are further assumed to be independent of each other. Also, the parameter S_t represents the underlying state at time t and takes on the value of 1 if x_t is in the first regime and 2 if x_t is in the second regime. In this context, the value of S_t depends on the corresponding wind direction of x_t.

The transition probabilities between different states are given by $p_{12} = P(S_t = 2|S_{t-1} = 1)$ and $p_{21} = P(S_t = 1|S_{t-1} = 2)$ with the restrictions that $0 < p_{12} < 1$ and $0 < p_{21} < 1$ and the sum of each row of the corresponding transition matrix be equal to 1. Equation (2.8) is not identifiable for data modeling purposes because there is no unique way to identify the states and the two sub-equations are interchangeable. To avoid this, we introduce constraints on the constant terms c_1 and c_2 by requiring that $c_2 > c_1$. As will be shown later in the case study, we then uniquely define State 1 to be the easterly regime and State 2 the westerly regime since the average wind speed in the westerly regime is higher than the easterly regime. The underlying Markov process is assumed to be homogeneous for parameter estimation purposes.

2.4.2 Parameter estimation

To estimate the parameters in equation (2.8), we employ a modified version of the Markov Chain Monte Carlo (MCMC) method as illustrated by McCulloch and Tsay (1994). First, we need to construct prior distributions for the parameters and subsequently derive the conditional posterior distributions from these constructed prior distributions. Then, using these posterior distributions, we simulate possible parameter values using the Gibbs sampler. In our case, we use conjugate priors for simplicity. In the original paper, the underlying states are assumed to be unknown and are estimated simultaneously with the rest of the parameters. However, since wind directions are used to define the states, they are known to us for estimation purposes. Therefore, we omit the steps taken to simulate the transition probabilities using Beta distributions. Here, we adopt the convention that $f(w|\cdot)$ represents the conditional posterior distribution of w given all the rest of the parameters and the wind speed data.

We start off by constructing χ^2 prior distributions to simulate the values for the variances of the innovation series $\{\epsilon_{1t}\}$ and $\{\epsilon_{2t}\}$, which take the form of

$$f(\sigma_j^2) \sim \frac{v_j \lambda_j}{\chi^2_{v_j}} \qquad j = 1, 2, \tag{2.9}$$

and each corresponds to the easterly and westerly regimes respectively. Furthermore, let v_1 and v_2 represent the degrees of freedom for the innovation series in the easterly and westerly regimes. Also, the λ_j's are arbitrary constants. Let $x_t^{(1)} = (x_{\tau_1}, \ldots, x_{\tau_t})$ where $t = 1, 2, \ldots, n_1$ and $x_t^{(2)} = (x_{\omega_1}, \ldots, x_{\omega_t})$ where $t = 1, 2, \ldots, n_2$. Here, $x_t^{(1)}$ and $x_t^{(2)}$ represent historical observations for the easterly and westerly regimes respectively, with n_1 and n_2 representing the number of observations for each regime. Then, the corresponding conditional posterior distribution is

$$f(\sigma_j^2|\cdot) \sim \frac{v_j \lambda_j + R_j^2}{\chi^2_{v_j+n_j-p}} \qquad j = 1, 2, \tag{2.10}$$

where R_j or residual terms in equation (2.10) for both regimes can be calculated by

$$R_j = x_t^{(j)} - c_j - \sum_{i=1}^{p} \phi_{j,i} x_{t-i}^{(j)} \qquad j = 1, 2. \tag{2.11}$$

Here, the initial values for constant terms c_1 and c_2 can be found by taking the average of the easterly and westerly wind speed observations respectively. Subsequent values are generated from a normal posterior distribution. In addition, the initial values for $\phi_{1,i}$ and $\phi_{2,i}$ can be found by first fitting a pth order autoregressive process to the easterly and westerly wind speed observations individually and using the $\boldsymbol{\phi} := \{\phi_{1,i}, \phi_{2,i}\}$ estimates obtained by Maximum Likelihood estimation as initial values. Subsequent values are then generated from a multivariate normal posterior distribution. The details about prior and posterior constructions for the constant and $\boldsymbol{\phi}$ parameters are given below.

We first illustrate the construction of prior distributions for c_1 and c_2. Here we assume that $f(c_j) \sim N(\mu_{c_{0,j}}, \sigma^2_{c0,j})$ for $j = 1, 2$ depending on the underlying state. As mentioned above, the mean $\mu_{c_{0,j}}$ for $j = 1, 2$ are estimated using the sample mean of easterly and westerly wind speed observations. Furthermore, the variance $\sigma^2_{c_{0,j}}$ for $j = 1, 2$ can be estimated using sample variance of easterly and westerly wind speed observations. The corresponding conditional posterior distributions are

$$f(c_j|\cdot) \sim \mathrm{N}(\mu_{c_j}, \sigma^2_{c_j}), \tag{2.12}$$

where the variance and mean terms are given by

$$\sigma^2_{c_j} = \frac{\sigma^2_j \sigma^2_{c_{0,j}}}{n_j \sigma^2_{c_{0,j}} + \sigma^2_j} \qquad \mu_{c_j} = \sigma^2_{c_j} \left(\frac{n_j \overline{y}_{c_j}}{\sigma^2_j} + \frac{\mu_{c_{0,j}}}{\sigma^2_{c_{0,j}}} \right) \qquad j = 1, 2. \tag{2.13}$$

The values for σ^2_j can be simulated from equation (2.10). To get the values for $\overline{y}_{c_j}$, we need to first subtract the $\boldsymbol{\phi}$ terms in equation (2.8) from $x_t^{(j)}$ for each regime as follows:

$$y_t^{(j)} = x_t^{(j)} - \sum_{i=1}^{p} \phi_{j,i} x_{t-i}^{(j)}, \tag{2.14}$$

where the term $\overline{y}_{c_j}$ in equation (2.13) is then the sample mean of the calculated $y_t^{(j)}$ for $j = 1, 2$ respectively. Again, the initial values for the $\boldsymbol{\phi}$ parameters can be found by fitting an autoregressive process to the easterly and westerly wind speed observations and use the calculated parameter estimates as initial values. Subsequent values are simulated from a multivariate normal distribution that will be described in detail below.

We then move on to the estimation step for the $\boldsymbol{\phi}$ parameters in (2.8). We first construct the prior distributions for $\phi_{1,i}$ and $\phi_{2,i}$ for $i = 1, 2, \ldots, p$. Here we assume that these two groups of parameters follow a multivariate normal distribution respectively for simplicity. Let $\boldsymbol{\phi}_1 = (\phi_{1,1}, \ldots, \phi_{1,p})'$ and $\boldsymbol{\phi}_2 = (\phi_{2,1}, \ldots, \phi_{2,p})'$, then

$$f(\boldsymbol{\phi}_j) \sim \mathrm{MVN}(\boldsymbol{\mu}_{s_{0,j}}, \boldsymbol{\Sigma}_{s_{0,j}}) \qquad j = 1, 2. \tag{2.15}$$

Here, the mean vector is $\boldsymbol{\mu}_{s_{0,j}}$ and the covariance matrix is $\boldsymbol{\Sigma}_{s_{0,j}}$. As mentioned above, the values for elements in $\boldsymbol{\mu}_{s_{0,j}}$ are obtained using the Maximum Likelihood estimates by fitting a pth order autoregressive process to the easterly and westerly observations. To obtain the initial values for $\boldsymbol{\Sigma}_{s_{0,j}}$ covariance matrix, we use the idea of diffusion initialization by setting the diagonal elements to be a very large number (in our case, 1×10^6) with the rest of the elements being set to 0. The conditional posterior distribution can be shown to take the form of

$$f(\boldsymbol{\phi}_j | \cdot) \sim \mathrm{MVN}(\boldsymbol{\mu}_{s_j}, \boldsymbol{\Sigma}_{s_j}), \tag{2.16}$$

where $\boldsymbol{\mu}_{s_j}$ and $\boldsymbol{\Sigma}_{s_j}$ are given by

$$\boldsymbol{\Sigma}_{s_j} = \left(\frac{\sum_{t=1}^{n_j} \boldsymbol{X}_t^{(j)'} \boldsymbol{X}_t^{(j)}}{\sigma_j^2} + \boldsymbol{\Sigma}_{s_{0,j}}^{-1} \right)^{-1}, \tag{2.17}$$

$$\boldsymbol{\mu}_{s_j} = \boldsymbol{\Sigma}_{s_j} \left(\frac{\sum_{t=1}^{n_j} \boldsymbol{X}_t^{(j)'} \boldsymbol{X}_t^{(j)}}{\sigma_j^2} \widehat{\boldsymbol{\rho}}_j + \boldsymbol{\Sigma}_{s_{0,j}}^{-1} \boldsymbol{\mu}_{s_{0,j}} \right) \tag{2.18}$$

for $j = 1, 2$. Here, $\boldsymbol{X}_t^{(j)} = (x_{t-p+1}^{(j)}, x_{t-p}^{(j)}, \ldots, x_{t-1}^{(j)}, x_t^{(j)})'$ with $t = 1, 2, \ldots, n_j$ and $\widehat{\boldsymbol{\rho}}_j = (\widehat{\rho}_{j1}, \widehat{\rho}_{j2}, \ldots, \widehat{\rho}_{jp})'$. For the elements in $\boldsymbol{X}_t^{(j)}$, we set $x_k^{(j)} = 0$ if $k < t$. To get the estimates of $\widehat{\boldsymbol{\rho}}_j$, we need to subtract the constant terms from the wind speed observations $x_t^{(j)}$ for each regime, i.e.,

$$w_t^{(j)} = x_t^{(j)} - c_j \qquad j = 1, 2. \tag{2.19}$$

We then regress this transformed series $w_t^{(j)}$ on $\boldsymbol{X}_t^{(j)}$ to obtain the least squares estimates for $\widehat{\boldsymbol{\rho}}_j$. Let $\boldsymbol{X}^{(j)} = (\boldsymbol{X}_{n_j}^{(j)}, \boldsymbol{X}_{n_j-1}^{(j)}, \ldots, \boldsymbol{X}_1^{(j)})$ and $\boldsymbol{w}^{(j)} = (w_1^{(j)}, w_2^{(j)}, \ldots, w_{n_j}^{(j)})'$, then the least squares estimates are given by

$$\widehat{\boldsymbol{\rho}}_j = (\boldsymbol{X}^{(j)'} \boldsymbol{X}^{(j)})^{-1} \boldsymbol{X}^{(j)} \boldsymbol{w}^{(j)} \qquad j = 1, 2. \tag{2.20}$$

Thus, we then proceed to draw the Gibbs sample from the conditional posterior distributions of (2.10), (2.12) and (2.16). The actual implementation is illustrated below:

1. We simulate the variance estimates σ_1^2 and σ_2^2 using (2.10).

2. These variance estimates, σ_1^2 and σ_2^2 simulated in step 1 will then be substituted into equation (2.16) to further simulate the values of the $\boldsymbol{\phi}$ parameters.

3. The variance estimates simulated in step 1 and the $\boldsymbol{\phi}$ parameters simulated in step 2 will then be substituted into equation (2.12) to further simulate values for the constant terms, c_1 and c_2.

4. We repeat step 1 to step 3 for $m + n$ times to obtain $m + n$ simulated values or Gibbs samples for the estimated parameters. We discard the first m random samples and keep the last n samples for parameter inferences. For example, we take the average of the n simulated values for each parameter to obtain point estimates and construct 90% credible intervals for each parameter by taking the 5th and 95th percentiles of the n random samples. These m samples are called burn-in samples and are used to ensure that the Gibbs samples will closely approximate a random sample drawn from the joint posterior distribution of the parameters.

The $m+n$ Gibbs samples can be shown to form a Markov chain (Geman and Geman, 1984). If this corresponding Markov chain is irreducible and ergodic, then the Gibbs samples will converge weakly to a stationary distribution which takes the form of the joint posterior distribution of all the parameters. The MCMC method is a Bayesian approach to estimate Markov regime switching models. It provides us greater flexibility in choosing different prior and posterior distributions and also in adjusting the hyperparameters involved. In addition, since parameters are treated as random variables, it is also possible to observe their distribution and see how the values of these parameters vary from one iteration to another.

2.4.3 State evolution and probabilistic forecasts

In the framework of the Markov regime switching model, to predict future values, we need to first predict the underlying states that the system would take in the future, since the states determine which autoregressive process will be used for prediction. In this paper, we will use a modified version of the Markovian local bootstrap (MLB) to generate possible series of states or paths that the system would evolve, by creating bootstrap samples from the underlying Markov process. We modified some steps in the original paper by Paparoditis and Politis (2001, 2002) using ideas from circular statistics to take into account the fact that the underlying states (wind directions) are angular variables.

For simplicity, we assume that the underlying Markov process is of order 1. Here, wind directions are all measured in degrees. Let θ_T be the observation at time T the last time point. In the general case, our aim is to simulate possible states until future time point $T + h$ where h represents the forecast horizon. Given historical wind direction data $\theta_1, \theta_2, \ldots, \theta_T$, the circular MLB proceeds as follows:

1. To obtain a bootstrapped value for θ^*_{T+1}, choose a value randomly from the set $\{\theta_{s+1} : \min(|\theta_T - \theta_s|, 360^\circ - |\theta_T - \theta_s|) \leq d \text{ and } 1 \leq s < T - 1\}$, where d is a bandwidth parameter. Assume that this chosen value is θ_{k_1}.

2. For the next bootstrapped value θ^*_{T+2}, choose a value randomly from the set $\{\theta_{s+1} : \min(|\theta_{k_1} - \theta_s|, 360^\circ - |\theta_{k_1} - \theta_s|) \leq d \text{ and } 1 \leq s < T - 1\}$. Suppose this value chosen is θ_{k_2}.

3. Repeat Steps 1 and 2 until forecast horizon h for N times to get a matrix of generated wind direction values, $\mathbf{\Theta}^* = (\!(\theta^*_{i,T+j})\!)$, for $i = 1, 2, \ldots, N$ and $j = 1, \ldots, h$. Here, each column represents the possible directions that the system would take at that time point. For example, the first column represents all possible directions the system would take at time $T + 1$ and so on.

4. From these simulated directions $\mathbf{\Theta}^*$, we convert them to the underlying states by defining $S^*_{i,T+j} = 1$ to be all $\theta^*_{i,T+j}$ such that $0^\circ < \theta^*_{i,T+j} \leq 90^\circ$ and $270^\circ < \theta^*_{i,T+j} < 360^\circ$. Also, define $S^*_{i,T+j} = 2$ to be all the degrees $\theta^*_{i,T+j}$ such that $90^\circ < \theta^*_{i,T+j} \leq 270^\circ$ where $i = 1, 2, \ldots, N$ and $j = 1, 2, \ldots, h$. Thus, we get a matrix of generated states $\boldsymbol{S}^* = (\!(S^*_{i,T+j})\!)$ for $i = 1, 2, \ldots, N$ and $j = 1, \ldots, h$.

5. With this matrix, $\boldsymbol{S}^*$ we then attempt to summarize all these possible states into just one simulated path. To do this, we count how many easterly or westerly states at each column. For example, if there are more easterly states or westerly

states in columns $S^*_{i,T+k}$ for $k = 1, 2, \ldots, h$, set $S_{T+k} = 1$ or $S_{T+k} = 2$ to get one single generated path $S = (S_{T+1}, \ldots, S_{T+h})$.

As shown in the original paper, the MLB is capable of producing bootstrapped samples that closely mimics the original Markov process. The drawing of samples from the sets defined above is actually a non-parametric way to estimate the transition densities. We could have skip Step 5 in the circular MLB algorithm and go ahead to produce probabilistic forecasts using sieve bootstrap for each generated states. However, we find that we get better performance in terms of coverage and root mean square error when we summarize all the generated states into one most probable state and proceed to use this state for wind speed forecasting. Therefore, with the most probable sample path S generated, we will use the sieve bootstrap as introduced by Alonso et al. (2002) to create fully probabilistic forecast distribution. This method provides us a way to generate a predictive distribution without having to make any parametric distribution assumptions and hence is applicable to a wide range of data. A detailed implementation of the sieve bootstrapping method is given below.

From equation (2.8), we calculate the residuals by $r_t^{(j)} = x_t - \widehat{c}_j - \sum_{i=1}^{p} \widehat{\phi}_{j,i} x_{t-i}$ for $t = p+1, \ldots, T$ and $j = 1, 2$ respectively. The estimated parameters $\widehat{c}_j$ and $\widehat{\phi}_{j1}, \ldots, \widehat{\phi}_{jp}$ are calculated using the MCMC method described in Section 3. We then create bootstrap samples $r_1^{(j)*}, \ldots, r_{T+h}^{(j)*}$ by sampling $T + h$ times with replacement from $\{r_t^{(j)}\}$ for $t = p+1, \ldots, T$ where $j = 1, 2$. Denote $x_T^*(h)$ to be the bootstrapped point forecast of the wind speed data at time T with a forecast horizon h. Then by referring to the elements in S, we can decide which autoregressive process in equation (2.8) will be used to forecast future values. In particular, if $S_{t+h} = j$, for $j = 1, 2$, then

$$x_T^*(h) = \widehat{c}_j + \widehat{\phi}_{j1} x_T^*(h-1) + \ldots + \widehat{\phi}_{jp} x_T^*(h-p) + r_{T+h}^{(j)*}. \tag{2.21}$$

If $h \leq 0$ then we set $x_T^*(h) = x_{t+h}$, which is the actual wind speed data observed. Below, we summarize the steps taken for the entire estimation and forecasting procedures as elaborated in this section,

1. Conduct preliminary investigations to study the structure of the data and to determine the order of the autoregressive process using the ACF and partial ACF plots.

2. Fit the model as in equation (2.8) to the data and use the modified MCMC method for parameter estimation.

3. Generate the most probable path of the underlying state, S using the circular MLB algorithm as described.

4. For this generated path, we repeat the sieve bootstrapping procedure for B times to get B values for $x_T^*(h)$.

Thus, from these B forecasts, we can construct predictive distributions and conduct inferences. For example, we construct 90% prediction intervals by calculating the 5th and 95th percentiles of these B forecasts.

3 Case study with real life data

In this case study, our aim is to produce 2-hour ahead probabilistic forecasts of wind speed at Vansycle. Here, the proposed BMRS model uses only Vansycle wind speed and direction data as oppose to the spatial-temporal models of RSTD and TDD which use both Vansycle's wind speed data and off-site data from Kennewick and Goodnoe Hills. For the wind data at Vansycle, we will concentrate on data collected in the year 2008 (8784 data points), which is the most recent available. We will first divide the data into 2 sets, the first set or the training set is used for estimation purposes and the second set is used for forecasting evaluation purposes. Referring to the paper by Gneiting et al. (2006), we adopt the sliding window method for parameter estimation of 45 days (1080 hours). In particular, the initial training set starts from the 24th hour November 16, 2007 to the 23th hour December 31, 2007, while the evaluation set is the entire year of 2008. The 24th hour of December 31, 2007 is not added because we are forecasting 2 steps ahead. To illustrate the wind dynamic differences between months in the evaluation set, performance measures are calculated separately for each month. Thus, the first 45 days of data will be used to produce probabilistic forecasts using the algorithm described in the previous section, and also to calculate performance measures by comparing these forecasts with observations in the evaluation set. Then, data points in the evaluation set will be moved into the estimation set with the first value in the estimation set discarded, keeping the size of the window constant at 1080 data points. This process is repeated until data points in the evaluation set are exhausted. Missing data are handled by substituting the corresponding previous year's value.

Before we begin actually implementing the proposed method, we first investigate the structure and properties of the wind speed and direction data using data collected in the first 10 months of 2007. We start by studying the behavior of the wind direction data at Vansycle. Figure 2 shows the distribution of wind direction at Vansycle against time starting from January to October 2007.

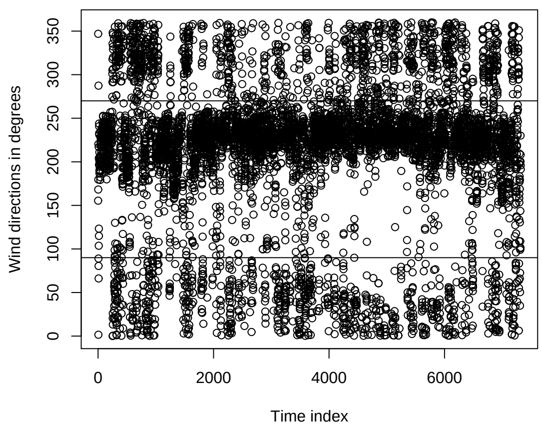


Figure 2: Wind direction at Vansycle in degrees from Jan 1st to Oct 30th 2007.

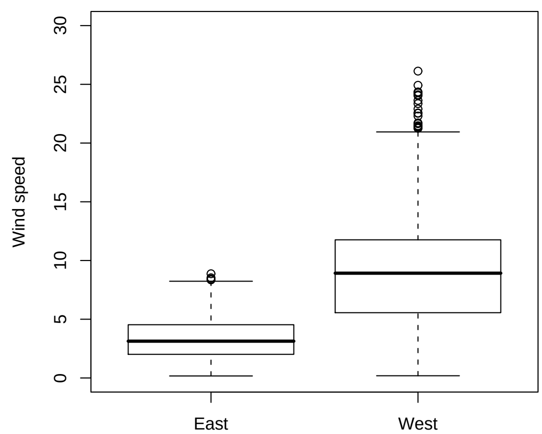


Figure 3: East and West wind speeds at Vansycle from Jan 1st to Oct 30th 2007.

In this plot, we observe that most points fall within the region from 90 degrees to 270 degrees (marked by two horizontal lines), while the rest of the points are scattered outside that region, and appears to be concentrated around 0 or 360 degrees. Here by definition, the first region corresponds to the westerly regime and the second region is the easterly regime. The high concentration of points in the westerly regime indicates that westerly winds are more frequent, while observations in the easterly regime are much more dispersed. This alternating wind directions from west to east suggest that we can model the wind speed observations using two distinct regimes. Furthermore, by separating Vansycle's wind speed observations according to the easterly and westerly regimes respectively, and plotting their box plots shown in Figure 3, we see that there is a separation of the regimes where wind speeds in the westerly regime are higher than their eastern counterpart. This suggests that westerly winds are driven by a different underlying wind dynamics then their eastern counterpart, and implies that we should fit different model dynamics to the westerly and easterly wind speed data. We then fit Vansycle's wind speed data using the proposed BMRS model. As illustrated in the previous section, we model the two different wind dynamics of Vansycle's wind speed data by fitting autoregressive processes of the same order to both regimes, while at the same time we assume that these dynamics are driven by a first order two state Markov chain with the states defined by the corresponding wind directions. To determine the order of the AR processes, we investigated the behavior of Vansycle's wind speed observations by plotting the corresponding time series plot shown in Figure 4. For comparison purposes, we also included wind speed data from Kennewick and Goodnoe Hills. The time series plots show one week wind speed observations at these 3 locations starting from May 1st, 2007. Taking hint from Gneiting et al. (2006) and also by referring to the plots, we notice that the wind speed data at Vansycle (and also the other 2 wind towers) appears to show diurnal trend. This fact is confirmed by observing the periodogram of Vansycle's data, which shows a peak at frequency around 1/24 as illustrated in Figure 5. We remove this trend using two pairs of sine and cosine functions in time. This diurnal de-trending step is applied to the evaluation set and as the forecast procedure moves forward, the coefficients of the trigonometric functions will be reestimated and updated. We then experiment with different autoregressive (AR) models and find that the 4th order AR process is the lowest order AR process that provides an adequate fit for Vansycle's de-trended wind speed data.

Model diagnostics and the Ljung-Box statistics plots show that the residuals obtained from fitting an AR(4) model are white noise and uncorrelated. Thus, based on these preliminary investigations, we decide to fit a different AR(4) process for each regime. For the parameter estimation step, we experiment empirically with several combinations of the possible values for v_j, λ_j for $j = 1, 2$ in equation (2.9). We find that by setting all $v_j = 3, \lambda_j = 0.8$ for $j = 1, 2$, we get a good coverage probability for the constructed prediction intervals. In addition, we further experiment with various values of m and n to ensure that the Gibbs random sample will converge to the final

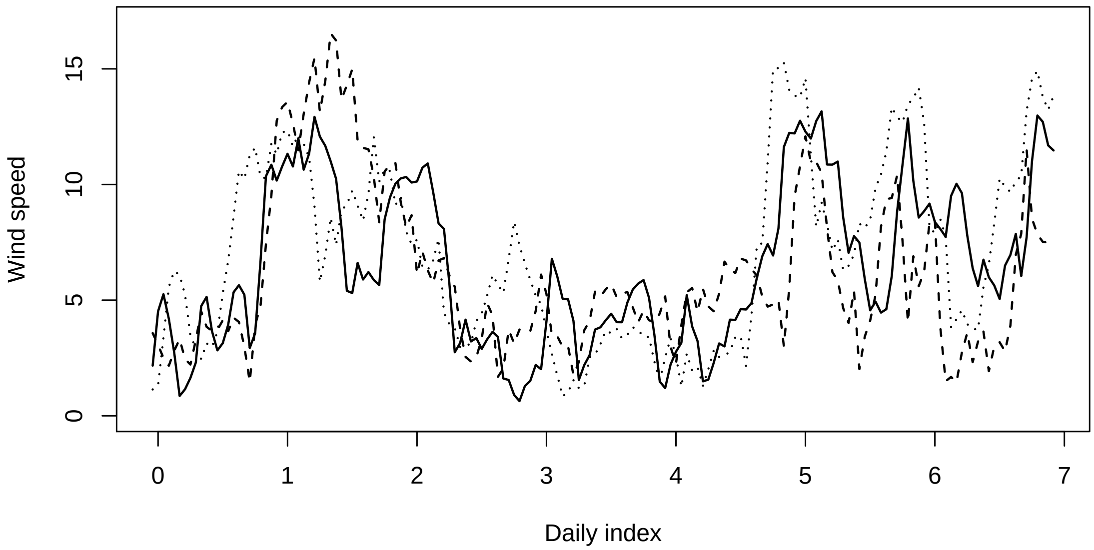


Figure 4: Time series plot for Vansycle (solid line), Kennewick (dashed line) and Goodnoe Hills (dotted line) wind speed for the first week of May 2007.

joint posterior distribution of the parameters. In our study, we set m and n to be 1000. In the circular MLB algorithm, we again experiment empirically to find the optimal value of the bandwidth, d and find that setting d from 0 to 7 will result in good performance measures. In our case, we set d to be 5. Also, we set N to be 50 and B to be 2000. Increasing these two values will only result in marginal improvement in the forecasts performance and we chose these values for computational efficiency. For the forecasting step, we set the forecast horizon, h to be 2 since our aim is to produce 2-hour ahead probabilistic forecasts.

3.1 Probabilistic forecast assessment

We employ the root mean square error (RMSE) as a performance measure to assess point forecasts. To calculate RMSE for our model, we take the mean of the B bootstrapped forecasts as point forecasts. In addition, we calculate the Mean Absolute Error (MAE) as an alternative measure to RMSE. In this case, we use the median of the B bootstrapped forecasts as point forecasts. Besides that, we use coverage

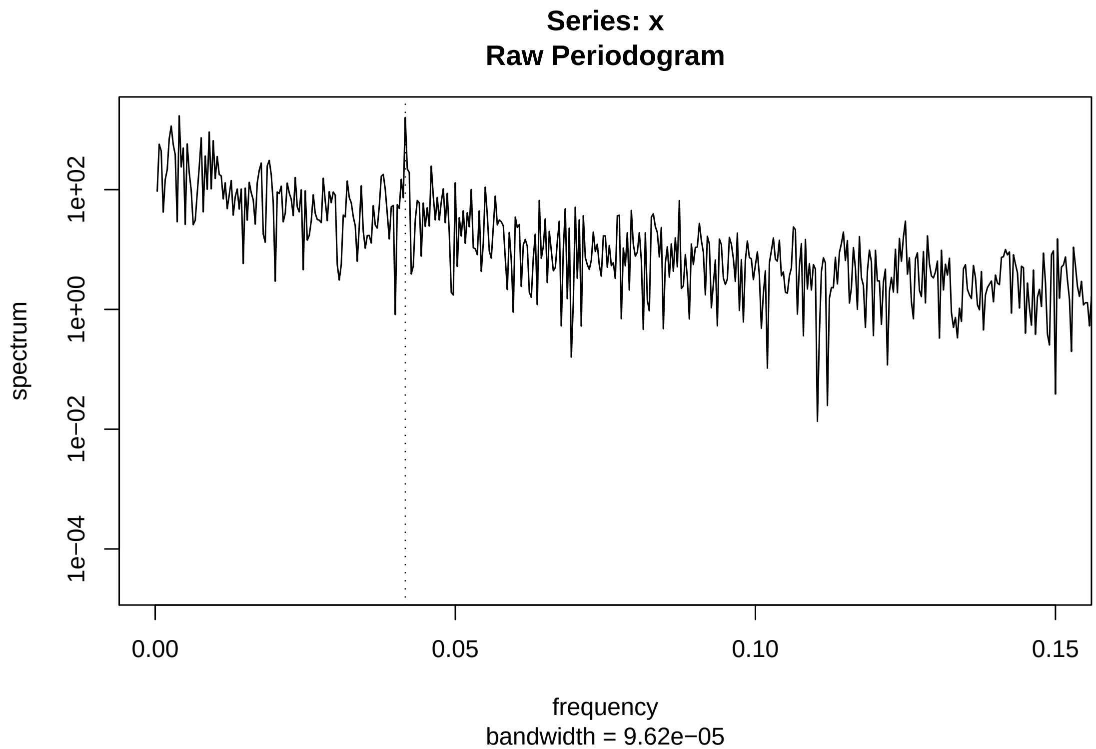


Figure 5: Periodogram for Vansycle wind speed data from Jan 1st to Oct 30th 2007. Vertical line shows peak frequency at 1/24 or 0.041667 Hz.

probability and confidence interval length as measures to assess the constructed prediction intervals. Generally, the $100(1-\alpha)\%$ prediction interval in the BMRS model can be found by letting $Q_{\alpha/2}$ be the lower $\alpha/2$ quantile and $Q_{1-\alpha/2}$ be the upper $\alpha/2$ quantile for the B bootstrapped forecasts. Then, to get the coverage probability, we calculate the proportion of times observations in the valuation set fall within $Q_{\alpha/2}$ and $Q_{1-\alpha/2}$. To calculate the length, we simply just take $Q_{1-\alpha/2}-Q_{\alpha/2}$. In this paper, we concentrate on getting 90% prediction intervals by setting $\alpha = 0.5$. To summarize the performance of the probabilistic forecast distribution, we employ Continuous Rank Probability Score (CRPS) measure introduced by Gneiting et al. (2005). In our case, we use an empirical version of the CRPS. Let f_i be one of possible forecast values generated from equation (2.21) where $i = 1, 2, \ldots, B$, and obs_t be the actual observation in the evaluation set at time t for $t = 1, 2, \ldots, q$, with q being the size of the evaluation set. Then the empirical CRPS at each evaluation time point t is given by

$$CRPS_t = \frac{1}{B}\sum_{i=1}^{B}|f_i - obs_t| - \frac{1}{2B^2}\sum_{i=1}^{B}\sum_{j=1}^{B}|f_i - f_j|\,. \tag{3.1}$$

The final CRPS score can be obtained by averaging $CRPS_t$ over the evaluation set. Besides calculating the CRPS, we also construct rank histogram to assess how well the bootstrapped probabilistic forecasts represent the true uncertainty of the wind observations. We proceed by first ordering the B bootstrapped forecasts in increasing order to create $B+1$ bins. We keep track of which bin observations from the evaluation set fall into, and assign rank i to a observation if it falls in bin i for $i = 1, \ldots, B+1$. We can then construct histogram based on these q rank values. In the ideal case, each bootstrapped forecasts represent an equally likely scenario, thus observation in the evaluation set is equally likely to fall between any 2 bootstrapped forecasts. In other words, flat histogram is preferable. They are other possible possibilities: the histogram is dome-shaped if the bootstrapped forecasts are over-dispersed where most observations fall in the middle, it is U-shaped if the forecasts are under-dispersed where most observations fall outside or near the extremes.

We shall see how our proposed model (BMRS), which utilizes only on-site wind speed and direction data, perform against the various measures discussed above. Besides applying the BMRS model to Vansycle, we are also interested to see how this model perform when applied to a new location. Thus in addition to Vansycle, we refit BMRS to wind speed and direction data of Kennewick and Goodnoe Hills. Here, the BMRS model has the advantage of the ease of transition from one location to another since it uses only on-site data.

Tabel 1 shows performance measures calculated for Vansycle when the evaluation period is from January to December 2008. Tables 2 and 3 show results for Kennewick and Goodnoe Hills respectively. Here, we calculate the performance measures for each month to see how these measures vary from month to month for BMRS. The overall measure in the last column are calculated by treating the entre year 2008 as a continuous evaluation period without breaking the evaluation period month by month. All together, there are 8784 two-hour ahead hourly forecasts. The target coverage for this study is taken to be 90%.

Table 1: Forecast performance measures calculated for the Vansycle data. RMSE: Root mean square error, MAE: Mean absolute error, CP: Coverage probability(90%), CLEN: Coverage length, CRPS: Continuous Ranked Probability Score

Measure	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Overall
RMSE	2.52	2.32	2.39	2.06	1.91	2.31	1.99	2.28	1.69	1.97	2.43	2.35	2.20
MAE	1.82	1.69	1.82	1.57	1.41	1.64	1.54	1.70	1.29	1.48	1.76	1.73	1.62
CP	90.46%	89.80%	88.71%	91.53%	91.67%	86.39%	89.52%	87.37%	93.47%	87.50%	84.86%	88.04%	88.95%
CLEN	7.85	7.78	7.41	7.47	6.56	6.53	6.68	6.35	6.51	6.02	6.76	7.11	6.92
CRPS	1.35	1.27	1.32	1.15	1.04	1.23	1.12	1.23	0.94	1.08	1.32	1.28	1.19

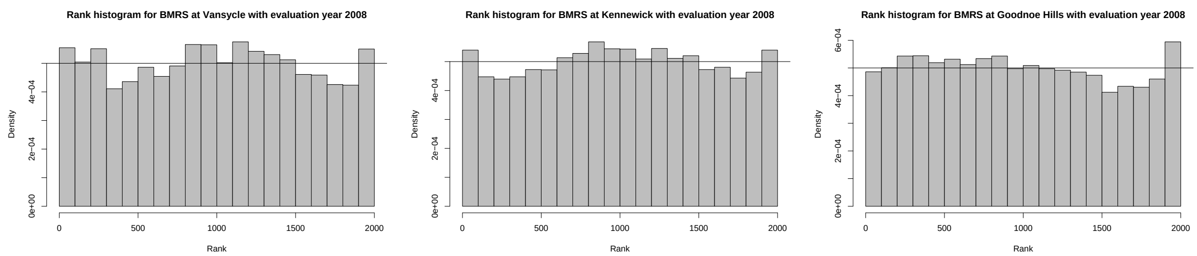
Table 2: Same measures calculated for the Kennewick data.

Measure	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Overall
RMSE	2.58	2.50	2.60	2.54	2.15	2.59	2.24	2.24	1.78	2.41	2.38	2.51	2.39
MAE	1.99	1.80	1.95	1.87	1.59	1.89	1.76	1.68	1.34	1.71	1.71	1.81	1.76
CP	88.84%	89.94%	85.89%	91.25%	93.28%	85.56%	90.05%	87.10%	93.19%	86.29%	89.72%	87.37%	89.20%
CLEN	8.33	8.28	7.84	8.27	7.84	7.24	7.58	6.95	6.63	6.50	7.33	7.23	7.50
CRPS	1.43	1.34	1.43	1.38	1.17	1.39	1.26	1.24	0.98	1.28	1.26	1.33	1.29

Table 3: Same measures calculated for the Goodnoe Hills data.

Measure	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Overall
RMSE	2.04	2.15	2.18	1.91	1.92	1.91	1.69	1.83	1.66	1.81	2.00	2.15	1.94
MAE	1.55	1.57	1.64	1.48	1.46	1.46	1.31	1.39	1.22	1.40	1.47	1.51	1.45
CP	89.65%	90.37%	88.44%	90.42%	91.40%	89.31%	89.92%	86.29%	90.69%	87.63%	88.33%	87.77%	89.21%
CLEN	6.56	6.63	6.83	6.75	6.68	6.32	5.68	5.26	5.35	5.48	5.79	5.85	6.10
CRPS	1.12	1.15	1.19	1.07	1.06	1.06	0.94	1.01	0.89	1.01	1.08	1.13	1.06

Figures 6 show the rank histograms for the Vansycle, Kennewick and Goodnoe Hills data for BMRS. Instead of showing the histograms for each month, we constructed histograms by treating the year 2008 as a continuous evaluation period for each tower to illustrate the quality of the forecasts in a concise way. In these plots, horizontal line represents the height of the histogram under the ideal situation, i.e. perfect spread.



(a) Vansycle (b) Kennewick (c) Goodnoe Hills

Figure 6: Rank histograms for BMRS model applied to Vansycle, Kennewick, and Goodnoe Hills with evaluation year 2008.

Tables 1–3 together show that the BMRS prediction intervals have coverage that are very close to the nominal 90% level. This is further reinforced in Figure 6, where the rank histogram is close to uniform, indicating that the bootstrapped probabilistic forecasts are able to capture the uncertainty in the data very well. We can conclude that BMRS produces wind speed forecasts that are sharp and well calibrated. This is true despite the fact that we have not assumed any distributional assumptions on the true errors.

4 Conclusion

Essentially, the proposed model is based on the belief that the prediction of a quantity of interest can be improved by incorporating appropriate auxiliary variables, which in this case is the wind direction at Vansycle. In this paper, we model the dynamics of wind speed data at Vansycle using the Markov Regime Switching model where we fit 4th order autoregressive processes to each regime respectively. In addition, we introduce the circular Markovian local bootstrap method to generate the most probable state that the system would take in the future. To arrive at the final ensemble probabilistic forecasts, we sieve bootstrap historical residuals and apply the recursive autoregressive prediction formulas. There are several ways we can further improve the BMRS model. As a first step, we might consider incorporating off-site wind observations, if they are available from the vicinity of Vansycle into the proposed model. In this context, we might include current or past wind speed series from Kennewick and Goodnoe Hills as additional regressors in equation (2.8). Research has shown that these type of spatial-temporal modeling approach will help improve the predictive performance. For example, Alexiadis et al. (1999) show that forecasts of wind speed and wind power at Thessaloniki Bay, Greece can be improved by incorporating off-site observations. Also, Gneiting et al. (2006) show that the inclusion of off-site observations as additional predictors in their RSTD model produces more accurate forecasts than traditional univariate time series models. However, as mentioned in the introduction, the lack of reliable off-site wind speed observations in many developing areas in the world will often force us to utilize only on-site data. On top of that, the introduction of these predictors will make the resulting forecast computation much more intensive. In addition, we are currently investigating possible extension of the circular Markovian local bootstrap method to deal with cases when the underlying Markov process is of order 2 or higher. This proposed method provides a way to create distribution-less probabilistic forecasts of wind speed using only on-site wind speed and direction data. As we can distinguish different wind direction regimes in many parts of the world, it is hoped that the proposed method can be used widely for wind speed forecasting, with applications in wind energy generation and management. This proliferation of improved forecasting technique will help make wind power a more viable choice for an alternative fuel, which would ultimately supplement and complement our existing fossil-based energy source.

5 Acknowledgement

The authors want to thank the Bonneville Power Administration and the Oregon State University Energy Resources Research Laboratory for their generosity in providing the wind speed and direction data.

References

Abraham, B. and Ledolter, J. (2005). *Statistical Methods for Forecasting.* Wiley, Hoboken, NJ.

Alexiadis, M. C., Dokopoulos, P. S., and Sashamanoglou, H. S. (1999). Wind speed and power forecasting based on spatial correlation models. *IEEE Transactions on Energy Conversion*, 14(3):836–842.

Alonso, A. M., Peña, D., and Romo, J. (2002). Forecasting time series with sieve bootstrap. *Journal of Statistical Planning and Inference*, 100(1):1–11.

Bao, L., Gneiting, T., Raftery, A. E., Grimit, E. P., and Guttorp, P. (2009). Bias correction and bayesian model averaging for ensemble forecasts of surface wind direction. Techical Report 557, Department of Statistics, University of Washington.

Brown, B. G., Katz, R. W., and Murphy, A. H. (1984). Time series models to simulate and forecast wind speed and wind power. *Journal of Climate and Applied Meteorology*, 23:1184–1195.

Campbell, S. D. and Diebold, F. X. (2005). Weather forecasting for weather derivatives. *Journal of the American Statistical Association*, 100(469):6–16.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics*, 7(1):1–26.

Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distributions and the bayesian restoration of images. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741.

Gneiting, T., Larson, K., Westrick, K., Genton, M. G., and Aldrich, E. (2006). Calibrated probabilistic forecasting at the stateline wind energy center: The regime-switching space-time method. *Journal of the American Statistical Association*, 101(475):968–979.

Gneiting, T., Raftery, A. E., Westveld, A. H., and Goldmann, T. (2005). Calibrated probabilistic forecasting using ensemble model output statistics and minimum crps estimation. *Monthly Weather Review*, 133:1098–1118.

Hamilton, J. D. (1988). Rational expectations econometrics analysis of changes in regime: An investigation of the term structure of interest rates. *Journal of Economic Dynamics and Control*, 12:385–423.

Hamilton, J. D. (1989). A new approach to the economic analysis of non-stationary time series and the business cycle. *Econometrica*, 57(2):357–384.

Haslett, J. and Raftery, A. E. (1989). Space-time modeling with long-memory dependence: Assessing ireland's wind power resource. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 38(1):1–50.

Hering, A. S. and Genton, M. G. (2010). Powering up with space-time wind forecasting. *Journal of the American Statistical Association*, 105(489):92–104.

McCulloch, R. E. and Tsay, R. S. (1994). Statistical analysis of economics time series via markov switching models. *Journal of Time Series Analysis*, 15:523–539.

Paparoditis, E. and Politis, D. N. (2001). A markovian local resampling scheme for nonparametric estimators in time series analysis. *Econometric Theory*, 17(3):540–566.

Paparoditis, E. and Politis, D. N. (2002). The local bootstrap for markov processes. *Journal of Statistical Planning and Inference*, 108(1–2).

Politis, D. N. (2003). The impact of bootstrap methods on time series analysis. *Statistical Science*, 18(2):219–230.

Psaradakis, Z. (1998). Bootstrap-based evaluation of markov-switching time series models. *Econometric Reviews*, 17(3):275–288.

Sloughter, J. M., Gneiting, T., and Raftery, A. E. (2010). Probabilistic wind speed forecasting using ensembles and bayesian model averaging. *Journal of the American Statistical Association*, 105(489):25–35.

T. Gneiting, T., Stanberry, L. I., Grimit, E. P., Held, L., and Johnson, N. A. (2008). Assessing probabilistic forecasts of multivariate quantities, with applications to ensemble predictions of surface winds. *Test*, 17:211–235.

Thorarinsdottir, T. L. and Gneiting, T. (2010). Probabilistic forecasts of wind speed: Ensemble model output statistics using heteroskedastic censored regression. *Journal of the Royal Statistical Society Series A (Statistics in Society)*, 173:371–388.

Tsay, R. S. (2005). *Analysis of Financial Time Series*. Wiley, Hoboken, NJ.